FACCE-MACSUR

A framework for assessing the uncertainty in crop model predictions

Daniel Wallach[1]*, Mike Rivington[2]

[1] * INRA, Toulouse, France
daniel.wallach@toulouse.inra.fr
[2] The James Hutton Institute, Craigiebuckler, Aberdeen. AB15 8QH. UK.
mike.rivington@hutton.ac.uk

Deliverable type: Report
File name: D-C4.1.2_Framework for uncertainty analysis
Deliverable reference num.: D-C4.1.2

Instrument:	Joint Programming Initiative
Topic:	Agriculture, Food Security, and Climate Change
Project:	Modelling European Agriculture with Climate Change for Food Security (FACCE-MACSUR)
Start date of project:	1 June 2012
Duration:	36 months
Theme, Work Package:	CropM, WP 4
Deliverable reference num.:	D-C4.1.2
Deliverable lead partner:	INRA
Due date of deliverable:	month 12
Submission date:	2014-07-04
Confidential till	-

Revision	Changes	Date
1.0	First Release	2014-07-04

Abstract/Executive summary

It is of major importance in modeling to understand and quantify the uncertainty in model predictions, both in order to know how much confidence to have in those predictions, and as a first step toward model improvement. Here we show that there are basically three different approaches to evaluating uncertainty, and we explain the advantages and drawbacks of each. This is a necessary first step toward developing protocols for evaluation of uncertainty and so obtaining a clearer picture of the reliability of crop models.

Table of contents

Introduction

Crop models, like all mathematical models of complex systems, are only approximations to the real world. The predictions of these models are never perfect; they have errors compared to observed responses. We speak of "uncertainty" in model predictions, to represent the fact that simulated behavior is not identical to observed behavior. Until the observation is actually made we have only simulated values, so we are uncertain as to what the observed behavior will be.

It is of major importance in modeling to understand and quantify the uncertainty in model predictions, both in order to know how much confidence to have in those predictions, and as a first step toward model improvement. The purpose of this paper is to present a framework for quantifying and understanding the uncertainty in model predictions.

The definition of uncertainty

We begin be defining model error ε:

$$\varepsilon_i = y_i - f(X_i;\hat{\theta}) \quad (1)$$

where the index i refers to some particular plot, y_i is the observed value of the response in question for that plot (e.g. observed yield) and $f(X_i;\hat{\theta})$ is the simulated value of the response, which depends on the explanatory variables for that field X_i (often daily weather, soil characteristics, management and initial conditions) and the parameter values of the model $\hat{\theta}$.

The plot i can be any plot in our target population. Each plot will have a different error ε_i and those errors will have some distribution. We refer to the distribution of ε_i as the model predictive uncertainty. Often rather than dealing with the full probability distribution, we estimate just model mean squared error of prediction.

Estimating the distribution of model error

In standard regression theory we can obtain an expression for the distribution of model error (Myers 2007). However, that is not applicable to the crop model situation, because standard regression analysis is based on several assumptions which probably don't apply to crop models. First, standard regression analysis (Wallach 2011). We must therefore estimate the uncertainty in crop model predictions without simple assumptions about model error. Three different general approaches are available, as detailed below.

a. Use hindcasts

The basic assumption here is that errors of prediction in the future will be similar to errors that were observed when using the model in the past. Then if we have hindcasts for a model, we can use the probability distribution function (pdf) of the errors in those hindcasts as our estimate of the probability distribution function for model error in the future.

This assumption will be justified if both the past and future conditions are random samples from the same underlying population. This approach estimates uncertainty for the model which has some specific parameter values.

There have been numerous studies where error for individual models has been evaluated using historic data (for example Abedinpoura et al. 2012; Ben Nouna et al. 2000; Soler et al. 2007; Mkhabela & Bullock 2012; Biernath et al. 2011).

The advantage of basing uncertainty on error in hindcasts is that it automatically takes into account all sources of error. It is important to keep in mind that if past data has been used to estimate model parameters (often referred to as model calibration or model tuning), then hindcast errors for those same data are no longer representative of future uncertainty.

b. Propagate uncertainty in parameters or inputs

The general procedure here is to posit a probability distribution for the uncertain factors (model parameters or input variables), run the model multiple times with different values drawn from that distribution, and then take the resulting distribution of outcomes as a representation of the uncertainty resulting from the uncertainty in those factors. Here we treat $\hat{\theta}$ as a random variable. The distribution for each individual parameter often comes from the literature, and represents the range of values for the parameter that have been reported. Examples can be found in (Richter et al. 2010; Dzotsi et al. 2013).
If a statistical procedure to parameter estimation is used, then the statistical procedure in general provides information about parameter uncertainty and also about model residual error. (Iizumi et al. 2009; Wallach et al. 2012) for example estimated crop model parameters, and their distribution, using a Bayesian approach.
One can also study the uncertainty due to uncertainty in model inputs, i.e. in the components of X. Examples are given in Aggarwal 1995, Mearns et al 1999, Rivington 2006, Roux et al. 2014; Ruane et al. (2013).

c. Estimate uncertainty based on multi-model ensembles

Multiple research groups around the world have developed different crop models that can be used for predicting the same outcomes. Rather than selecting a single model and accepting its results, the approach here is to take the distribution of simulated values as an approximation to the uncertainty in the predictions.
The model is here treated as chosen randomly from an ensemble of models. The uncertainty is that due to the random choice of a model from within the ensemble of models. Examples are (Asseng et al. 2013; Palosuo et al. 2011; Rötter et al. 2012; Bassu et al. 2014).
One advantage of this method is that it takes into account uncertainties in model structure and equations and also, at least to some extent, uncertainties in model parameters. Another advantage of this approach is that one can evaluate the uncertainty as a function of the situation (value of X). A final advantage is that it is well adapted to studying the uncertainties in cascades of models (Tao et al. 2009; Tao & Zhang 2010).
A major shortcoming of the multi-model approach is that there could well be cases where all the models agree quite closely but all the models are wrong. This would correspond to the case where the first term on the right in eq.
Fehler! Verweisquelle konnte nicht gefunden werden. is large, and the second term on the right, which is taken to represent uncertainty, is small.

Open questions

There are still a number of open questions related to approximating the uncertainty in predictions of crop models. First, despite the fact that there have been multiple studies on model uncertainty, we still do not have a clear overall picture of how large prediction uncertainty is, for different situations, for example for different outputs, for different methods of calibration, etc. Partly this is because there has been no thorough review of the literature on this topic. Such a review is difficult, because of the diversity of methods for evaluating uncertainty that have been used. Hopefully, the recent increase in collaboration among modeling groups will lead to more homogeneous studies of model uncertainty and to a better understanding of how reliable crop models are.
There are also questions related to the methodology of estimation of prediction uncertainty. Estimating the uncertainty due to parameter uncertainty is one major difficulty. In particular, there is no commonly accepted definition of the reference value for the parameters, and no commonly accepted method for quantifying the uncertainty in estimating the reference value.

The use of model ensembles also poses methodological problems. One set of questions relates to the choice of the models in the ensemble. How should these be chosen? How many are needed? Another question concerns the combination of ensemble modeling and propagation of parameter uncertainty. In climate modeling, there are studies where one uses both multiple models and multiple parameter values for each model (Rosenzweig et al., 2013b; Nelson et al., 2014).. One could envision a similar procedure for crop models, but this does not seem to have been done. We do not know if this would substantially increase the estimated uncertainty, and make our estimations more realistic.

References

Abedinpoura, M. et al., 2012. Performance evaluation of AquaCrop model for maize crop in a semi-arid environment. *Agricultural Water Management*, 110, pp.55–66.

Asseng, S. et al., 2013. Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, 3(9), pp.827–832.

Bassu, S. et al., 2014. How do various maize crop models vary in their responses to climate change factors? *Global change biology*, 20(7), pp.2301–20.

Biernath, C. et al., 2011. Evaluating the ability of four crop models to predict different environmental impacts on spring wheat grown in open-top chambers. *European Journal of Agronomy*, 35(2), pp.71–82.

Dzotsi, K.A., Basso, B. & Jones, J.W., 2013. Development, uncertainty and sensitivity analysis of the simple SALUS crop model in DSSAT. *Ecological Modelling*, 260, pp.62–76.

Iizumi, T., Yokozawa, M. & Nishimori, M., 2009. Parameter estimation and uncertainty analysis of a large-scale crop model for paddy rice: Application of a Bayesian approach. *Agricultural and Forest Meteorology*, 149(2), pp.333–348.

Mkhabela, M.S. & Bullock, P.R., 2012. Performance of the FAO AquaCrop model for wheat grain yield and soil moisture simulation in Western Canada. *Agricultural Water Management*, 110, pp.16–24.

Myers, R.H., 2007. *Classical and modern regression with applications*, Boston : PWS-Kent.

Ben Nouna, B., Katerji, N. & Mastrorilli, M., 2000. Using the CERES-Maize model in a semi-arid mediterranean environment. Evaluation of model performance. *European Journal of Agronomy*, 13, pp.309–322.

Palosuo, T. et al., 2011. Simulation of winter wheat yield and its variability in different climates of Europe: A comparison of eight crop growth models. *European Journal of Agronomy*, 35(3), pp.103–114.

Richter, G.M. et al., 2010. Sensitivity analysis for a complex crop model applied to Durum wheat in the Mediterranean. *European Journal of Agronomy*, 32(2), pp.127–136.

Rötter, R.P. et al., 2012. Simulation of spring barley yield in different climatic zones of Northern and Central Europe: A comparison of nine crop models. *Field Crops Research*, 133, pp.23–36.

Roux, S., Brun, F. & Wallach, D., 2014. Combining input uncertainty and residual error in crop model predictions: A case study on vineyards. *European Journal of Agronomy*, 52(Part B), pp.191–197.

Soler, C.M.T., Sentelhas, P.C. & Hoogenboom, G., 2007. Application of the CSM-CERES-Maize model for planting date evaluation and yield forecasting for maize grown off-season in a subtropical environment. *European Journal of Agronomy*, 27(2), pp.165–177.

Tao, F. et al., 2009. Modelling the impacts of weather and climate variability on crop productivity over a large area: A new super-ensemble-based probabilistic projection. *Agricultural and Forest Meteorology*, 149(8), pp.1266–1278.

Tao, F. & Zhang, Z., 2010. Adaptation of maize production to climate change in North China Plain: Quantify the relative contributions of adaptation options. *European Journal of Agronomy*, 33(2), pp.103–116.

Wallach, D. et al., 2012. Assessing the uncertainty when using a model to compare irrigation strategies. *Agronomy Journal*, 104, pp.1274–1283.

Wallach, D., 2011. Crop model calibration: A statistical perspective. *Agronomy Journal*, 103, pp.1144–1151.